Project: Diamond Prices

Complete each section. When you are ready, save your file as a PDF document and submit it here: https://classroom.udacity.com/nanodegrees/nd008/parts/235a5408-0604-4871-8433-a6d670e37bbf/project#

Step 1: Understanding the Model

Answer the following questions:

1. According to the model, if a diamond is 1 carat heavier than another with the same cut, how much more should I expect to pay? Why?

   Assuming the same cut and clarity level for the diamond, a diamond that is 1 carat heavier would cost 8,413 more. This is the coefficient to the "Carat" variable in the regression model.

2. If you were interested in a 1.5 carat diamond with a **Very Good** cut (represented by a 3 in the model) and a **VS2** clarity rating (represented by a 5 in the model), how much would the model predict you should pay for it?

   The regression model provided is:
   Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity.

   Following the new diamond specifications given,
   Price=- 5,269 + 8,413 x 1.5 + 158.1 x 3 + 454 x 5 = **10,097.8**

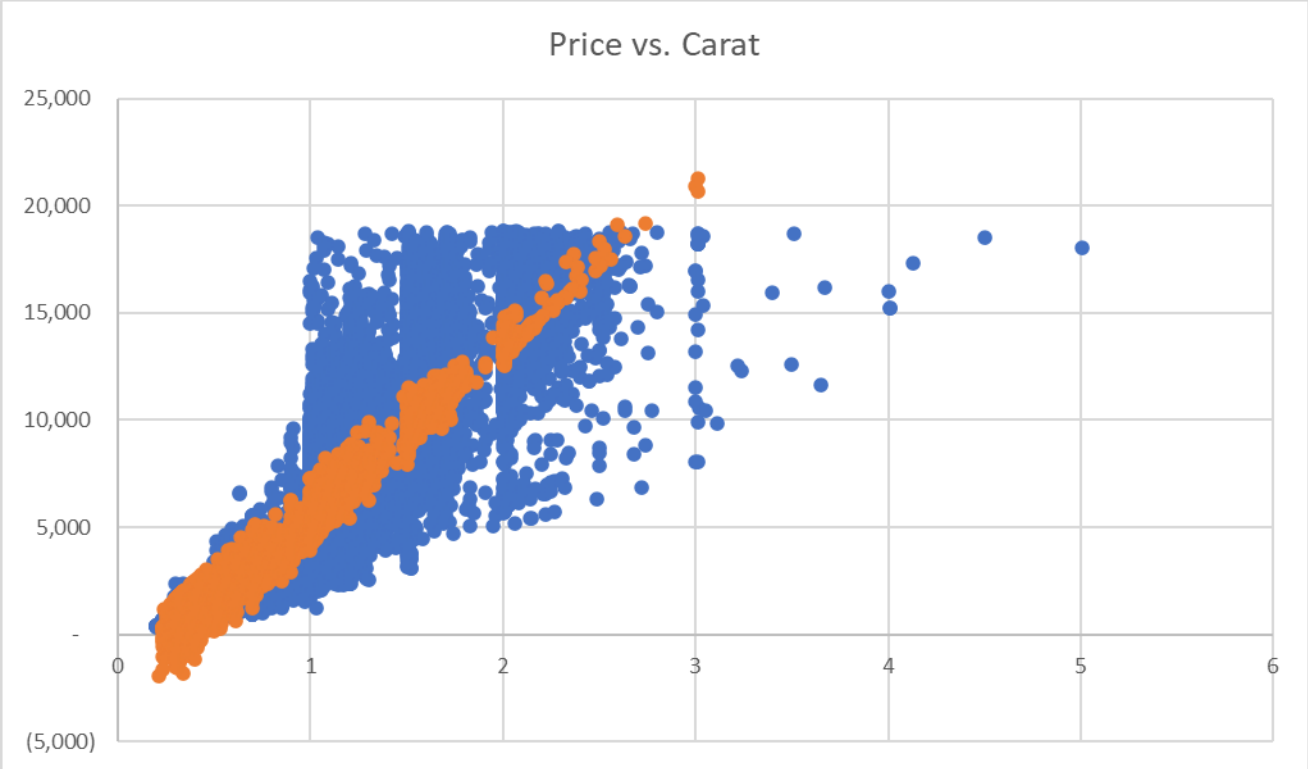
   Thus, the model would predict that we must pay 10,097.8 for this specified diamond.

Step 2: Visualize the Data

Make sure to plot and include the visualizations in this report. For example, you can create graphs in Excel and copy and paste the graphs into this Word document.

1. Plot 1 - Plot the data for the diamonds in the database, with carat on the x-axis and price on the y-axis.

2. Plot 2 - Plot the data for the diamonds for which you are predicting prices with carat on the x-axis and predicted price on the y-axis.
    - **Note**: You can also plot both sets of data on the same chart in different colors.
3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?



While the model-predicted prices follow the increasing trend of prices as carat increase, I initially did not feel that confident in its ability to predict prices for three reasons:

1. The model erroneously results to some negative prices for 291 diamonds. This is because we have a large negative intercept of 5,269 and a model heavily dependent on the carat level with a coefficient of 8,413.
2. In actual prices, there is a large jump in prices when 1 carat is breached. The model does not reflect this.
3. The model does not consider 'color' as a variable. Based on research, this appears to be an important factor.
4. The slope of our model also appears steeper than the actual prices.

Trying out other models will still result to some incorrect interpretations. For instance, changing the intercept to 0:

Price = 6,987 x Carat - 466 x Cut + 93.5 x Clarity, where -466 for a coefficient is counterintuitive. However, only 16 diamonds (from 291) will have prices below 0.

Trying another model considering the color level (1-7) and using a non-zero intercept again:

Price = -4,609 + 8,727 x Carat + 161 x Cut + 494.75 x Clarity - 303 x Color, which again has -303 for a coefficient and is counterintuitive. However, 296 diamonds will have a price below 0. Hence, the results for this are even worse than the original model.

Step 3: Make a Recommendation

Answer the following questions:

1. What price do you recommend the jewelry company to bid? Please explain how you arrived at that number.

The jewelry company can bid $8,213,466 for the whole set. This is equivalent to 70% of the resulting bid price from the original model. However, should the seller not agree, the company can bid up to $8,656,048 and this would still be acceptable considering the first model had 291 diamonds with negative prices.

The computations are below:

Model	Bid Price for Whole Set	Final Bid Price (70% of price)
Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity.	$11,733,523	$8,213,466
Price = 6,987 x Carat - 466 x Cut + 93.5 x Clarity	$12,365,783	$8,656,048